DOCUMENT RESUME

ED 284 903 TM 870 456

AUTHOR Zimowski, Michele F.; Wothke, Werner

TITLE Purification of Spatial Tests: An IRT Analysis of Spatial and Reasoning Components in "Spatial" Tests.

PUB DATE 22 Apr 87

NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).

PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Adults; *Aptitude Tests; *Content Validity; Factor Analysis; Intelligence Tests; *Item Analysis; Latent Trait Theory; Logical Thinking; Postsecondary Education; Problem Solving; Sex Differences; *Spatial Ability; Testing Programs; Test Interpretation; Test Items; Verbal Tests; Visualization; *Visual Measures

IDENTIFIERS BILOG Computer Program; *Johnson O Connor Aptitude Tests; *Mental Rotation Tests; Raven Progressive Matrices

ABSTRACT

Tests of spatial ability were analyzed for their analog (visuospatial) and nonanalog (verbal reasoning) components, using factor analyses of items and test scores. The self-selected sample consisted of over 2000 clients (average age about 26 or 27) employing the Johnson O'Connor Research Foundation's aptitude evaluation services in 12 metropolitan laboratories across the nation. Spatial tests administered were: (1) Incomplete Open Cubes; (2) Guilford-Zimmerman Spatial Visualization; (3) Wiggly Block; and (4) Paper Folding. Raven's Advanced Progressive Matrices was included as a measure of Spearman's general intelligence factor (G), and two tests of verbal reasoning ability (analytical and inductive) were given as a reference for nonanalog abilities. TESTFACT results distinguished item features that resisted nonanalog solution strategies (high loadings on first factor) from item features that promoted nonanalog strategies (high loadings on second factor). The sex difference favoring males depended on the analog component, and was largest for the Guilford-Zimmerman test, which requires the most "spatial" processing. The factor structure for test scores was less clear, partly because of ceiling effects. Two nonanalog factors were found, and it was noted that many tests currently used to evaluate spatial abilities depend largely on verbal reasoning components. (LPG)

Purification of Spatial Tests: An IRT Analysis of Spatial and Reasoning Components in "Spatial" Tests

Michele F. Zimowski, The University of Chicago and
Werner Wothke, Northwestern University

Spatial ability, measured by tests that require the mental manipulation of configural information, has been of considerable interest since Thorndike (1921) and McFarlane (1925) first demonstrated that it was relatively independent of Spearman's General Intelligence factor (G). Spatial measures were then routinely included in the multiple factor work of the 1920s and 30s (e.g., Kelley, 1928). In the studies reviewed by Wolfle (1940), the Spatial factor was second only to the Verbal factor in its frequency of occurrence. Additional support for the relative independence of these abilities soon appeared in the validation work of the 1930s, 40s, and 50s. Measures of spatial and verbal abilities exhibited distinct patterns of correlations with technical proficiences and academic success in various subject areas (See McGee, 1979, for a review). In the 1940s and 50s this growing evidence in favor of a distinct Spatial factor led to examinations of the factorial structure of the spatial domain (French, 1951; Guilford & Lacey, 1947; Thurstone, 1950). It was during this period and for these purposes that many of the spatial tests currently in use were orginally developed.

Although these tests all required the processing of visuospatial stimuli, not all measured an ability that was relatively distinct from verbal skills. As early as 1950, Spearman and Jones noted that items of visuospatial content could:

> be readily solved in two distinct manners. One may be called analytic, in the sense that attention wanders from one element of the figures to another. The other mode of operation is comparatively synthetic, in that the figures (or their constituents) . mentally grasped in much larger units (sometimes called "wholes"). The former procedure, not the latter, tends to load noegenetic [i.e., congeneric] processes with G (p.70).

Similar distinctions among "spatial" tests and processing modes have since appeared not only in the individual differences literature, but in the information processing literature as well (e.g., analytic versus holistic(analog) processing of visuospatial information, Metzler & Shepard, 1974, Cooper, 1976; analytic versus nonanalytic spatial ability, Maccoby & Jacklin, 1974; propositional versus spatial/imagery models of visuospatial representation and processing, Kosslyn & Shwartz, 1977, Paivio, 1977; nonanalog versus analog visuospatial tests, Zimowski, 1985, Zimowski & Wothke, 1986).

Despite this growing body of evidence suggesting that many spatial measures contain verbal components, the term "spatial" is still used rather indiscriminately in the individual differences literature to refer to any test that requires the processing of visuospatial information (e.g., Eliot & Smith, 1983; Caplan, MacPherson, & Tobin, 1985). As a result, conclusions drawn in this literature tend to be test-dependent. This is especially true of studies that have focused on identifying the biological and sociocultural determinants of individual and sex differences in spatial ability. Progress in this and other areas now depends on a better understanding of the item features that promote or require verbal reasoning solution strategies and a means for identifying relatively pure measures of spatial (analog) ability.

The work of Zimowski (1985) and Zimowski and Wothke (1986) is a step in this direction. In their review of item-feature effects, they identify item attributes associated with analog (spatial, holistic) and nonanalog (verbal reasoning) solution strategies. They find that analog items, items resisting nonanalog solution strategies and displaying analog effects, share a number of properties. First, they involve judgments among rotated stimuli. Other transformation tasks are less resistant to solution by nonanalog processes. Second, the stimuli differ by orientations other than 180 degrees. Because simple verbal rules such as "the right side now becomes the left side" can be used to solve 180-degree items, these items tend to have a nonanalog component. Third, the distractors of these tasks are mirror images of the reference stimuli or structurally equivalent forms. When mirror-image distractors are not used, the problems are readily solved through "feature-extraction" strategies, e.g., identification of incongruent portions of the figures. Fourth, the items require whole-whole rather than part-whole or part-part comparisons. Subjects report using serial comparison and other nonanalog strategies on items of the latter type. These items also produce effects consistent with a nonanalog model of information processing (see Pylyshyn, 1979). Fifth, analog items require the rotation of an entire object as a rigid whole rather than the rotation of only one or several pieces of the object relative to the whole. Finally, solution time restrictions are imposed on the items to inhibit solution through other than analog means. Almost any spatial item, even one with properties that resist nonanalog solution, can be solved through these means if enough time is allowed for their application. Zimowski and Wothke (1986) use this list of item features to classify existing instruments as relatively pure (analog) or impure (nonanalog) measures of spatial ability.

Because examinations of spatial item-feature effects are primarily found in the information processing literature, their classification scheme is largely based on studies that differ from psychometric evaluations of abilities in many important ways. First, variation in ability (the focus of the psychometric approach) is often deliberately minimized in these studies through extensive training or selection of subjects for homogeneous aptitude (see the work of Shepard & Metzler, 1971, for a case in point). Second, the studies are usually based on very small samples of subjects and the items are presented under laboratory conditions which differ markedly from standard testing procedures.

The objective of the present study is to further test the validity of the Zimowski-Wothke classification scheme in a psychometric evaluation of spatial tests. Through application of recent advances in IRT methodology, the present research extends the examination of item-feature effects to spatial test items that are administered to large samples under standard testing procedures.

Sample and Test Selection

The study was conducted in cooperation with Johnson O'Connor Research Foundation (JOCRF) with the practical goal of improving the Foundation's measurement of spatial ability. The self-selected sample consisted of clients employing the Foundation's aptitude evaluation services in twelve laboratories across the nation.

Two spatial tests (The Incomplete Open Cubes test, Zimowski, 1985; A modification of the Guilford-Zimmerman Spatial Visualization test, Bock & Kolakowski, 1973) and one measure designed to assess Spearman's G in a culture-free manner (The Advanced Progressive Matrices, Raven, 1962) were added to the Foundation's test battery for the study. They were administered along with the test battery under standard conditions by Foundation staff.

Several tests from the regular battery were selected for analysis. They include the Foundation's measures of spatial ability, The Wiggly Block (O'Connor, 1928) and Paper-Folding Test (French, Ekstrom, & Price, 1963), and two measures of verbal reasoning ability (Analytical Reasoning (AR) and Inductive Reasoning (IR)).

Not all participants were administered all of the tests; sample sizes for the individual tests ranged between 2199 and 2814. The IRT analyses were based on all available responses, while comparative analyses of the scaled scores were based on the subsample of participants who completed all tests. Complete measurements were available for 801 females and 917 males. These subgroups are comparable with respect to average age (27.6 and 26.3 years old, respectively) and years of education (14.4 and 14.1, respectively).

Description of the Spatial Tests

The spatial tests of the study are described below with reference to the classification scheme of Zimowski and Wothke (1986).

The Incomplete Open Cubes Test (IOC)

The Incomplete Open Cubes test (Zimowski, 1985) was especially constructed to measure analog and nonanalog processing in relatively distinct subsets of items. The version used in this study consists of 47 pairs of incomplete (parts of) cubes. The first three pairs are practice items, the remainder test items. The items are presented on slides and their exposure times are individually controlled. Subjects are asked to determine whether two incomplete parts of cubes fit together to form a complete open cube. An item from the test is shown in Figure 1a.

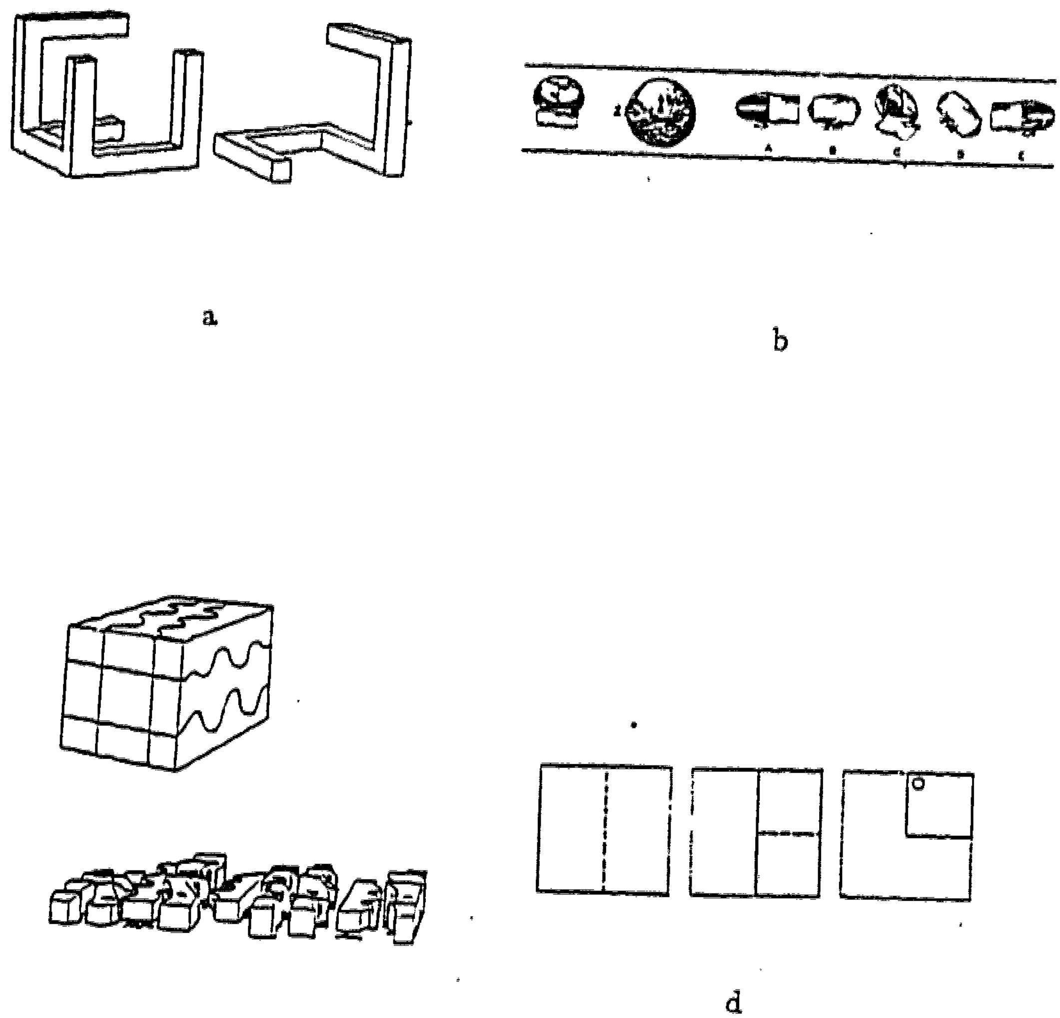

Figure 1: Examples of items from the Incomplete Open Cubes Test (a), The Guilford-Zimmerman Spatial Visualization test (b), The Wiggly Block (c), and the Paper Folding Test (d).

The items of this test are classified according to specific physical characteristics of the cubes (e.g., the distribution of segments between parts of the cubes) and general form. In the latter classification, cubes that fit together to form a complete cube are referred to as "compatible" cubes, those that cannot be joined together as "incompatible" cubes. Compatible cubes are further distinguished by the number of degrees (i.e., 45 or 90) that one, either one, of the compatible parts must be rotated in order to be joined with the other part. Incompatible cubes may be mirror images (MI) or nonmirror images (NMI). If either incomplete cube of a mirror image pair is replaced by its mirror image, the other left as is, the two cubes become compatible. The NMI condition refers to the lack of this relationship. Items in this condition were especially constructed to encourage nonanalog processing of the stimuli. They contain distinctive features that readily permit solution without rotation.

The Guilford-Zimmerman Spatial Visualization Test (GZ)

In the Guilford-Zimmerman Spatial Visualization test (Guilford & Zimmerman, 1947) subjects are asked to mentally rotate a picture of a clock in a specified direction and select the alternative that shows the clock in its final position (see Figure 1b). Each alternative is a picture of the clock as viewed from a different perspective. As a result, the alternatives do not contain distinctive features that allow for rapid elimination of incorrect alteratives. In the Bock-Kolakowski (1973) modification of the test used in this study, the items are presented on slides and their exposure times are individually controlled. This test is assumed to be a relatively pure measure of spatial ability.

The Wiggly Block (WB)

The Wiggly Block (O'Connor, 1928) differs from most other spatial tests because it is a performance measure. The test consists of wooden blocks that have been sawn into 4, 6, 9, or 12 pieces. The cuts are wavy or wiggly rather than plane. All of the pieces are approximately equal in volume and weight, but no two are identical. Subjects are presented with a random arrangement of the pieces and are asked to assemble the block as quickly as possible (see Figure 1c for a sample item). Scores on the test are based on the amount of time is takes to assemble each block.

The test is thought to have a small nonanalog component because the pieces of the blocks have distinctive features that provide information about their global location. Corner, outer-edge and inside-center pieces are readily distinguished by their number of flat surfaces. This information can be used to reduce the number of potentially compatible pieces that must be rotated for fit to any other piece. Its use should also reduce the amount of time required to assemble the blocks and, thus, improve performance on the test.

Paper-Folding Test (PF)

In each item of the paper-folding test (French, Ekstrom, & Price, 1963) the subject is shown a series of figures obtained from folding a square sheet of paper and punching a hole in the folded form. Each step of this folding and punching process is depicted in a separate figure of the series. The subject is asked to determine the position of the holes if the paper were unfolded. In the version in use at Johnson O'Connor the subject indicates the position of the holes on a square grid of paper.

Many of the items in the test can be readily solved through verbal rules. The item shown in Figure 1d, for example, can be solved through application of a symmetry principle.

Methodology

The TESTFACT implementation (Wilson, Wood, & Gibbons, 1984) of a full-information method for dichotomous item factor analysis (see Bock, Gibbons, & Muraki, 1985 for a detailed treatment of the method) was used to examine the factorial composition of several tests in the battery and to identify item-feature effects. The model is a multidimensional extension of a univariate ogive model to more than one dimension and is based on Thurstone's (1947) multiple factor formulation. The method provides estimates of the slopes, intercepts, standard difficulties, and factor loadings of the items.

A test of fit of the assumed factor model against a general multinomial alternative is provided by the chi-square approximation for the likelihood ratio test (G^2). This test is inaccurate in applications where the count of possible score patterns is much larger than the sample size. The difference in these statistics from alternative models is nonetheless distributed as a chi-square variable in large samples with degrees of freedom equal to the difference between those of the models. This change in G^2 as additional factors are added provides a test for the number of factors.

Once the appropriate model has been determined in this manner, expected a posteriori (EAP) estimates of a subject's ability on each factor may be computed on the basis of his item score pattern and the estimates of the factor loadings and standard difficulties.

Results

Full-information Item Factor Analyses

Separate full-information factor analyses were performed on the item response data from the IOC, GZ, and Raven's Advanced Progressive Matrices (PM) using the TESTFACT program. Preliminary one-dimensional item response analyses of these tests were performed with the BILOG program of Mislevy and Bock (1984) to determine whether a guessing model should be used in the factor analyses. Only the IOC showed evidence of guessing with a common asymtote value of .31. The average asymptote values for the GZ and PM were .03 and .04, respectively, too small to be of practical importance. Using the asymptote value obtained from BILOG, a guessing model was substituted for the normal response model in the TESTFACT analysis of the IOC.

In these applications of TESTFACT, there is some uncertainty associated with the statistical tests of the number of factors because the sample is self-selected into testing centers located in 12 metropolitan areas across the country. The effect is probably similar to that of cluster sampling which tends to inflate the values of the likelihood ratio chi-square statistics (see Zimowski & Bock, 1987). For this reason, the values of these statistics were adjusted to reflect an assumed design effect of 2.5.

The tests of fit for the IOC are presented in Table 1. As shown in this table, the chi-square value associated with the addition of the third factor is not significant, but the change in chi-square upon addition of the second factor is quite large. This change is roughly five times the corresponding change in degrees of freedom and clearly supports a two-factor model.

The varimax-rotated factor loadings from the two-factor solution are presented in Table 2. The first factor is primarily defined by the compatible and MI items, the second by the NMI and MI items. EAP estimates of the subjects' abilities were computed on the basis of these factor loadings for use in subsequent analyses.

TABLE 1

FULL INFORMATION FACTOR ANALYSIS TESTS OF FIT FOR THE INCOMPLETE OPEN CUBES TEST

Factor	Change G^2	df	p
2	226.20	43	$< .001$
3	43.23	42	.42

The tests of fit for the GZ are presented in Table 3. Two factors are apparently required to account for the item responses to this test but the patterns of factor loadings from the varimax and promax-rotated solutions are not readily associated with content similarities among the items. The promax-rotated factors are substantially intercorrelated ($r=.79$) and a large percent of the variance is attributable to the first principle factor (33.83) in comparison with the second (2.04). The change in chi-square upon addition of the second factor is small, roughly twice the change in degrees of freedom, in comparison with the change associated with the addition of the clearly defined second factor of the IOC.

Because of the uncertainty associated with these statistical tests in this study, and the lack of substantive evidence to support the two-factor solution, the GZ was assumed to be unidimensional. EAP estimates of ability based on a one-dimensional two-parameter model were calculated with

BILOG for use in subsequent analyses.

The tests of fit for the PM are presented in Table 4. Although, a two-factor model is indicated, neither the varimax nor the promax-rotated factor loadings are easily interpreted in terms of item content. The change in chi-square associated with the addition of the second factor is less than twice its degrees of freedom. Once again, the evidence does not strongly support a two-factor model. EAP estimates of ability based on a one-dimensional two-parameter model were therefore derived for use in subsequent analyses.

TABLE 2

IOC FACTOR LOADINGS FROM THE TWO-FACTOR VARIMAX SOLUTION

		Items with compatible cubes:		
Type	Item #	Analog	Non-Analog	Segments
	4	.27	.29	5/7
	13	.45	.34	5/7
	15	.40	.27	5/7
	18	.50	.25	5/7
	19	.37	.30	6/6
45°	29	.73	.03	4/8
	31	.46	.24	6/6
	36	.40	.24	5/7
	37	.70	.08	6/6
	39	.32	.23	5/7
	46	.64	.09	5/7
	7	.35	.25	5/7
	10	.73	.50	4/8
	11	.50	.31	4/8
	14	.40	.32	6/6
	23	.24	.21	5/7
90°	27	.40	.21	6/6
	30	.70	.07	5/7
	35	.67	-.03	5/7
	40	.46	.18	6/6
	41	.80	.22	4/8
	44	.32	.14	5/7

Items with incompatible cubes:

Type	Item #	Analog	Non-Analog	Distrib
MI	8	.55	.63	3/9
	9	.56	.66	4/8
	16	.45	.64	3/9
	21	.44	.70	3/9
	22	.26	.52	5/7
	24	.07	.44	6/6
	26	.04	.39	7/5
	28	.20	.50	6/6
	32	.23	.55	4/8
	38	.19	.54	4/8
	42	.13	.51	5/7
NMI	5	.20	.47	4/8
	6	.17	.67	3/9
	12	.10	.37	4/8
	17	.16	.75	3/9
	20	.13	.40	6/6
	25	.03	.46	4/8
	33	.26	.38	5/7
	34	.20	.43	6/6
	43	.31	.50	6/6
	45	.12	.51	5/7
	47	.19	.45	5/7

TABLE 3

FULL INFORMATION FACTOR ANALYSIS TESTS OF FIT FOR THE GUILFORD-ZIMMERMAN

Factor	Change G^2	df	p
2	69.02	29	$< .001$
3	35.42	28	.16

TABLE 4

FULL INFORMATION FACTOR ANALYSIS TESTS OF FIT FOR THE ADVANCED PROGRESSIVE MATRICES TEST

Factor	Change G^2	df	p
2	60.34	35	$< .005$
3	35.33	42	.40

Factor Analysis of the Test Scores

A maximum likelihood factor analysis was performed on the test scores to determine the factorial structure of the battery. EAP estimates of ability derived in the IRT analyses served as the test scores for the IOC, GZ, and PM in this analysis. The scores used by JOCRF in their evaluation of aptitude profiles served as the measures for the tests from their battery. These measures are raw scores based on reaction time.

Even a four-factor model failed to provide a good fit to the data ($G^2 = 8.91, df = 2, p = .01$). Scores from the second factor of the IOC and the Analytical Reasoning test had the lowest communalities, .40 and .38, respectively. The other measures exhibited communalities in excess of .50.

The varimax-rotated factor loadings from the four-factor model are presented in Table 5. The first factor is primarily defined by the analog factor of the IOC and the three other spatial tests. Raven's Advanced Progressive Matrices test exhibits the largest loading on the second factor, followed by the Paper-folding test. The Guilford-Zimmerman and the analytical reasoning tests also tend to load on this factor but to a lesser degree. The third factor is almost entirely defined by the two reasoning tests from the JOCRF battery. The fourth and final factor includes the nonanalog factor of the IOC, the Guilford-Zimmerman and the Paper-Folding test. The Wiggly Block also defines this factor, but to a smaller extent. In contrast with the other spatial tests, each of the IOC measures loads almost exclusively on just one factor. The promax-rotated factor loadings exhibit a similar overall pattern.

TABLE 5

FACTOR LOADINGS FROM THE ANALYSIS OF TEST SCORES
VARIMAX-ROTATED SOLUTION

Test	Factor 1	Factor 2	Factor 3	Factor 4
IOC-1	.76	.18	.10	.16
IOC-2	.17	.16	.11	.58
GZ	.49	.37	.12	.46
WB	.53	.26	.20	.34
PF	.47	.50	.13	.45
AR	.22	.35	.44	.14
IR	.06	.02	.86	.10
PM	.21	.65	.07	.19

Sex-differences

A sex differences favoring males on tests of spatial ability is frequently reported in the literature (e.g., Maccoby & Jacklin, 1974), but, as mentioned in the introduction, this finding tends to be test-dependent. For comparative purposes, the effect sizes associated with the male-female contrast in performance were computed for all tests in the battery. They are shown in Table 6 in ascending order of size.

TABLE 6

EFFECT SIZES FOR THE MALE-FEMALE CONTRAST

Test	Effect Size
GZ	.65
IOC-1	.48
WB	.44
IOC-2	.43
PF	.31
PM	.04
AR	-.05
IR	-.23

Discussion

The item factor decomposition of the Incomplete Open Cubes obtained in this study is in accord with the earlier findings of Zimowski (1985) and supports the distinction between item features and solution strategies first proposed by Zimowski (1985) and later developed by Zimowski and Wothke (1986). With few exceptions, items with features thought to resist nonanalog processing load on the first factor, while items with features thought to promote nonanalog processing load, almost exclusively, on the second factor.

The factors defined by these subsets of items are only slightly distinguished by the tendency of their factor scores to exhibit a sex difference favoring males. While the effect size observed for the analog factor of the IOC is .48, that for the nonanalog factor is .43. This result does not agree with the earlier work of Zimowski (1985) who found a large difference in these effect sizes. This failure to replicate is probably due to the effects of the self-selected sample. TESTFACT provides varimax-rotated factors that are uncorrelated in the population, but in this nonrandom sample the separation of the two factors is not complete. The scores from these factors correlate .26.

The pattern of effect sizes found for the other spatial tests in the study is, however, consistent with their classification as relatively pure (analog) or relatively impure (nonanalog) measures of spatial ability. The Guilford-Zimmerman test, which is assumed to be a relatively pure measure of spatial ability, shows the largest sex effect of all the measures in the study. The Wiggly Block exhibits a sex difference but to a lesser degree. The Paper Folding test, the least pure measure in the study, also displays the smallest sex difference. These results are in accord with the observation of Zimowski (1985) that the analog component of spatial tests is responsible for the sex difference favoring males.

The results from the factor analysis of the test scores are less clear. While the two spatial tests thought to have nonanalog components, the Wiggly Block and the Paper Folding test, load on the factors largely defined by the PM test and IOC-2, the Guilford-Zimmerman, which is classified as a relatively pure measure of spatial ability, also tends to load on these factors.

A possible explanation for the performance of the Guilford-Zimmerman lies in the ceiling effect observed on this test. The raw score distribution, shown in Figure 2, indicates that it was too easy for our self-selected sample. Many of the subjects were able to correctly answer all, or nearly all, of the 30 items in this test. It is possible that the standard time limit imposed on each item of the GZ was too generous for this group of verbally proficient individuals and allowed the successful application of nonanalog strategies. This interpretation explains the ceiling effect and the pattern of factor loadings, but it fails to account for the substantial sex effect found for this measure.

Another interpretation that is consistent with the analog classification of the Guilford-Zimmerman is that the ceiling effect obscured the factor pattern that would have otherwise been observed. The ceiling effect probably attentuated the sex difference as well, but apparently the effect was not large enough to produce a subtantial reduction in the size of the difference.

The emergence of two nonanalog factors in the factor analysis suggests that at least two, relatively distinct types of nonanalog strategies or abilities are used to solve spatial items. One of these abilities is represented by the Advanced Progressive Matrices and Analytical Reasoning. Both of these tests require an understanding of the logical relationships among elements and presumably measure logical reasoning. While the elements of the former are visuospatial stimuli, those of the latter are verbal terms. The second nonanalog factor, defined by the IOC-2, appears to represent a more specific ability to extract relevant distinctive features. This interpretation is supported by the loading of the Wiggly Block on this factor. The items of this test also contain distinctive features that can be used to bypass the rotation process (see the earlier description of the Wiggly Block).

The Paper Folding test exhibits loadings of similar magnitude on both factors suggesting that both types of abilities are used to solve items in this test. As illustrated earlier, verbal rules and logic can be used to solve many of the items in this test. The role of feature-extraction strategies is less clear. It is possible that the ability to identify the features of these items that permit solution through verbal rules is different from the ability to apply these rules.

In all, the results of this IRT-based study support the classification scheme of Zimowski and Wothke (1986). The inconsistencies found in this study are most likely due to the effects of nonrandom sampling. More generally, the study shows that the very feature that first distinguished spatial measures, their relative independence from verbal measures, is no longer characteristic of many of the "spatial" tests currently in use. If consensus is to be reached in substantive studies of spatial ability, workers must be aware of this fact when they tests for their studies. The classification scheme of Zimowski and Wothke provides useful guidelines for this selection process.

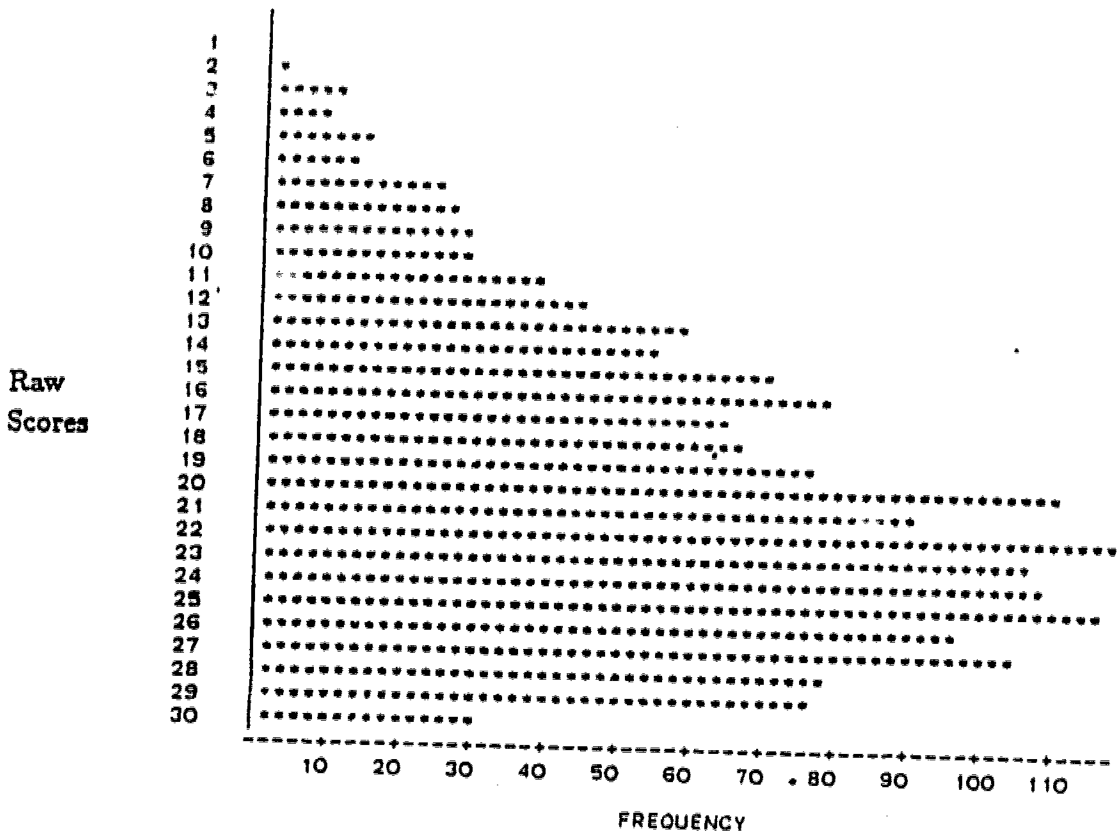

Figure 2: Raw score distribution from the The Guilford-Zimmerman Spatial Visualization test.

References

Bock, R.D., Gibbons, R.D., & Muraki, E. (1985). *Full-information factor analysis.* Chicago: National Opinion Research Center.

Bock, R.D., & Kolakowski, D. (1973). Further evidence of sex-linked major-gene influence on human spatial visualizing ability. *The American Journal of Human Genetics*, 25, 1-14.

Caplan, P.J., MacPherson, G.M., & Tobin, P. (1985). Do sex-related differences in spatial abilities exist?: A multilevel critique with new data. *American Psychologist*, 40, 786-799.

Cooper, L.A. (1976). Individual differences in visual comparison processes. *Perception and Psychophysics*, 19, 433-444.

Eliot, J.C., & Smith, I.M. (1983). *An international directory of spatial tests.* Windsor, England: NFER-Nelson.

French, J.W., Ekstrom, R.B., & Price, L.A. (1963). *Kit of reference tests for cognitive factors.* Princeton, NJ: Educational Testing Service.

French, J.W. (1951). The description of aptitude and achievement tests in terms of rotated factors. *Psychometric Monographs* (No. 5). London: Macmillan.

Guilford, J.P., & Lacey, J.I. [Eds.] (1947). *Printed classification tests.* Army Air Forces Aviation Psychology Program Research Report No. 5. Washington, DC: U.S. Government Printing Office.

Guilford, J.P., & Zimmerman, W.S. (1947). *The Guilford-Zimmerman aptitude survey – spatial visualization, form B.* Beverly Hills, CA: Sheridan Supply Company.

Kelley, T.L (1928). *Crossroads in the mind of man.* Stanford, CA: Stanford University Press.

Kosslyn, S. M., & Shwartz, S.P. (1977). A simulation of visual imagery. *Cognitive Science*, 1, 265-295.

Maccoby, E.E., & Jacklin, C.N. (1974). *The psychology of sex differences.* Stanford,CA: Stanford University Press.

McFarlane, M. (1925). A study of practical ability. *British Journal of Psychology, Monograph Supplement No. 8.*

McGee, M.G. (1979). *Human spatial abilities.* New York: Holt, Rinehart, and Winston.

Metzler, J., & Shepard, R.N. (1974). Transformational studies of the internal representation of three-dimensional structures. In Solso, R.L. (Ed.) *Theories in cognitive psychology: The Loyola symposium* (pp.147- 201). New York: Wiley.

Michael, W.B., Guilford, J.P., Fruchter, B., & Zimmerman, W.S. (1957). The description of spatial-visualization abilities. *Educational and Psychological Measurement*, **17**, 185-199.

Mislevy, R.J. & Bock, R.D. (1984). *BILOG: Maximum likelihood item analsis and test scoring.* Mooresville, Indiana: Scientific Software, Inc.

O'Connor, J. (1928). *Born that way.* Baltimore: Williams & Wilkins.

Paivio, A. (1977). Images, propositions, and knowledge. In J.M. Nicholas (Ed.), *Images, perception and knowledge.* Dordrecht, Holland: Reidel.

Pylyshyn, Z.W. (1979). The rate of "mental rotation" of images: A test of a holistic analogue hypothesis. *Memory and Cognition*, **7**, 19-28.

Raven, J.C. (1962). *Advanced Progessive Matrices, Set II* (1962 Revision). London: E.T. Heron & Co.

Shepard, R.N., & Metzler, J. (1971). Mental rotation of three- dimensional objects. *Science*, **171**, 701-703.

Spearman, C., & Jones, L.W. (1950). *Human ability.* London: Macmillan.

ΩThorndike, E.L. (1921). On the organization of intellect. *Psychological Review*, **28**, 141-151.

Thurstone, L.L. (1947). *Multiple factor analysis.* Chicago: The University of Chicago Press.

Thurstone, L.L. (1950). *Some primary abilities in visual thinking.* Chicago: The University of Chicago Psychometric Laboratory Report No. 59.

Wilson, D., Wood, R., & Gibbons, R.D. (1984). *TESTFACT: Test scoring and full-information item factor analysis.* Mooresville, Indiana: Scientific Software, Inc.

Wolfe, D. (1940). Factor analysis to 1940. *Psychometric Monograph No. 3.*

Zimowski, M.F. (1985). *Attributes of spatial test items that influence cognitive processing.* Unpublished doctoral dissertation, The University of Chicago, Department of Behavioral Sciences.

Zimowski, M.F. & Bock, R.D. (1987). *Full-information item factor analysis of test forms from the ASVAB CAT pool.* Chicago: National Opinion Research Center.

Zimowski, M.F. & Wothke, W. (1986). *The measurement of human variation in spatial visualizing ability: A process-oriented perspective.* Technical report No. 1986-1. New York: The Johnson O'Connor Research Foundation, Inc.